Journal of Multivariate Analysis **81**, 167–186 (2002)
doi:10.1006/jmva.2001.1998, available online at http://www.idealibrary.com on IDEAL®

R-estimation in Autoregression with Square-Integrable Score Function

Kanchan Mukherjee[1] and Z. D. Bai[2]

National University of Singapore, Singapore, Republic of Singapore
E-mail: stamk@nus.edu.sg, stabaizd@leonis.nus.edu.sg

Received April 25, 2000; published online November 13, 2001

This paper develops an asymptotic theory for R-estimation based on a square-integrable, not necessarily bounded, score function in the pth order stationary autoregressive model. Asymptotic uniform linearity of a class of linear rank statistics is established and the asymptotic normality of the corresponding R-estimators

weak convergence technique of Hájek, Jurečková and Koul, respectively.

AMS 1991 subject classifications: 62M10; 62G20.

Key words and phrases: R-estimation; autoregressive models; contiguity; robust estimation.

1. INTRODUCTION

Consider the pth order autoregressive model

$$X_i = \beta_1 X_{i-1} + \cdots + \beta_p X_{i-p} + \varepsilon_i, \qquad 1 \leqslant i \leqslant n, \tag{1.1}$$

where $\{X_i;\, 1-p \leqslant i \leqslant n\}$ are the observations, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ is the vector of unknown parameters (to be estimated) and $\{\varepsilon_i;\, 1 \leqslant i \leqslant n\}$ are error random variables satisfying the following.

(B.1) (Assumption on the error distribution): $\{\varepsilon_i;\, 1 \leqslant i \leqslant n\}$ are independent and identically distributed (i.i.d.) with density (pdf) f. Moreover, $E(\varepsilon_1) = 0$ and $Var(\varepsilon_1) = \sigma^2$, where $0 < \sigma < \infty$.

[1] The research is partly supported by the Academic Research Grant RP 3982708 from the National University of Singapore.

[2] The research is partly supported by the Academic Research Grant RP 3972712 from the National University of Singapore.

0047-259X/01 $35.00

(B.2) (Assumption on the autoregression parameter $\boldsymbol{\beta}$): The modulus of all roots of the pth degree polynomial $1-\beta_1 x-\cdots-\beta_p x^p$ are greater than unity.

Assumptions (B.1) and (B.2) ensure the existence of a stationary solution of (1.1) and we assume that $\{X_i; 1-p \leqslant i \leqslant n\}$ is a stationary sequence of random variables. See Brockwell and Davis (1996, Theorem 3.1.1) for more on this.

The commonly used methods of estimating $\boldsymbol{\beta}$ include the method of moments (Yule–Walker equation), the least squares and the maximum likelihood methods. Despite some of their attractive merits, these estimators are not very efficient when the error distributions are heavy-tailed. Alternative estimators, similar to those used in linear regression models, have been suggested by many researchers in the last two decades. For example, Denby and Martin (1979) and Koul (1986) proposed generalized M- and minimum distance estimators, respectively. Both Denby and Martin (1979) and Koul (1991) derived asymptotics of generalized M-estimators based on bounded score functions. Koul and Saleh (1995) considered autoregression quantiles and related class of L-estimators based on bounded score functions. All of these estimators are robust in terms of the asymptotic efficiency for heavy-tailed error distributions. The relative merits and demerits of these different estimators from the computational and the asymptotic efficiency points of view are similar to those of the analogous estimators for linear regression models and are discussed in detail in Koul (1992, Chapter 7).

There is a vast literature on the rank estimation (R-estimation) of parameters in linear regression models. Major contributions include Adichie (1967), Sen (1969), Jurečková (1971), Koul (1971), Jaeckel (1972) and Heiler and Willers (1988). R-estimators are sometimes preferable to their other competitors for their global robustness and efficiency considerations (classical Chernoff and Savage (1958) phenomenon). For details, see Hájek, Šidák and Sen (1999, Section 10.3) Koul (1992, Section 4.4) and Jurečková and Sen (1996, Section 3.4), among others. See also Remark 4.1 below.

Motivated by these, Koul (1992), Koul and Saleh (1993) and Koul and Ossiander (1994) developed the theory for R-estimation in autoregressive models by minimizing dispersions based on non-decreasing right-continuous bounded score functions of ranks, using the 'weak convergence technique'. Such technique based on weighted empirical processes works only for bounded score functions such as the one used for defining Wilcoxon type R-estimator. But, such technique does not work for unbounded score functions. However, R-estimation based on an unbounded square-integrable score function is also important from both theoretical and practical points of view because of the possibility of classical Chernoff and Savage (1958) phenomenon. Towards that, the R-estimator corresponding to the unbouded but square-integrable normal score function (van der Waerden type R-estimator) is conjectured to be asymptotically efficient at the Gaussian errors and is conjectured to outperform the least squares estimator for all other error densities. To date, the asymptotic normality result for the van der Waerden type R-estimator in the context of autoregressive setup is not available in the literature. In linear regression models, Jurečková (1971) developed the theory for R-estimation based on unbounded but square-integrable score function on (0, 1) using the "contiguity technique" and some stringent assumptions (later relaxed by Heiler and Willers (1988)) on the non-random regressors. Because of the dependence among the observations and the randomness of the regressors, this technique can not be invoked easily to obtain similar results in the autoregressive setup for R-estimators based on unbounded score functions. In this paper, we use a combination of Koul, Jurečková, and Heiler and Willers's techniques as well as some L^2-approximation techniques (of square-integrable functions by step functions) of Hájek to develop the theory for R-estimation in autoregression based on a square-integrable score function. Thus, this paper fills a gap in the existing literature on autoregression. In the process, we have also developed some interesting technical lemmas (proved in the Appendix), which are expected to be useful in further research. In particular, we remark that the asymptotic theory for generalized M-estimators in Denby and Martin (1979) and Koul (1991) and that of L-estimators in Koul and Saleh (1995) is developed based on bounded scores and it will be interesting to investigate whether the boundedness assumption can be relaxed using the techniques similar to what described in this paper.

The rest of this paper is organized as follows. A class of R-estimators based on square-integrable score functions is defined in Section 2. Section 3 states some technical results which are used heavily in Section 4. Section 4 discusses the asymptotic uniform linearity (AUL) of rank statistics and the asymptotic normality of R-estimators. The Appendix section contains the proofs of the technical results stated in Section 3.

2. R-ESTIMATOR

Let $\boldsymbol{Y}_{i-1} = (X_{i-1}, X_{i-2}, ..., X_{i-p})'$, $1 \leqslant i \leqslant n$, and $\bar{\boldsymbol{Y}} = \sum_{i=1}^{n} \boldsymbol{Y}_{i-1}/n$. Let φ: $(0, 1) \to \mathbb{R}^1$ be a (score) function belonging to the class

$$\mathscr{F} = \left\{ \varphi;\ \varphi\colon (0, 1) \to \mathbb{R}^1 \text{ is non-constant, non-decreasing and } \int_0^1 \varphi^2(u)\, du < \infty \right\}.$$

Define a linear rank statistic

$$\boldsymbol{S}_{\varphi}(\boldsymbol{t}) = n^{-1/2} \sum_{i=1}^{n} (\boldsymbol{Y}_{i-1} - \bar{\boldsymbol{Y}})\, \varphi\left(\frac{R_{it}}{n+1}\right), \qquad \boldsymbol{t} \in \mathbb{R}^{p},$$

where $R_{it} = \sum_{j=1}^{n} I(X_j - \boldsymbol{t}'\boldsymbol{Y}_{j-1} \leqslant X_i - \boldsymbol{t}'\boldsymbol{Y}_{i-1})$ (the $\boldsymbol{t}$-residual rank of the ith residual), $1 \leqslant i \leqslant n$. Also, let $R_{i\beta}$, the rank of ε_i among $\{\varepsilon_j; 1 \leqslant j \leqslant n\}$, be simply denoted by R_i. Following Koul (1992), an R-estimator of $\boldsymbol{\beta}$ corresponding to the score function φ is defined as

$$\hat{\boldsymbol{\beta}}_{\varphi} = \operatorname{argmin}\left\{ \sum_{j=1}^{p} |S_{\varphi j}(\boldsymbol{t})|;\, \boldsymbol{t} \in \mathbb{R}^{p} \right\},$$

where $S_{\varphi j}(\boldsymbol{t})$ is the jth coordinate of the vector $\boldsymbol{S}_{\varphi}(\boldsymbol{t})$, $1 \leqslant j \leqslant p$. For the existence of the solution to the above minimization problem and computation in the analogous setup, see Huber (1981, Section 7.3) and Koul (1992, Section 7.3b). Note also that this minimization problem may not always have unique solution. However, as in Jurečková (1971, Section 4) for the analogous case of linear regression models, it can be shown using AUL results that all solutions are asymptotically equivalent.

Remark 2.1. Strictly speaking, these estimators are not functions of the ranks of the t-residuals only. However, we borrow the terminology from the regression setting and still call them R-estimators for autoregression also. When, for example, $\varphi(u) = u - \frac{1}{2}$, $\hat{\boldsymbol{\beta}}_{\varphi}$ is an analogue of the Wilcoxon type R-estimator. When $\varphi(u) = \Phi^{-1}(u)$, where $\Phi(\cdot)$ is the standard normal distribution function, $\hat{\boldsymbol{\beta}}_{\varphi}$ is an analogue of the van der Waerden type R-estimator.

Remark 2.2. A theory of genuinely rank-based R-estimation can be developed, quite naturally, on the basis of serial rank statistic considered in a hypothesis testing context by Hallin, Ingenbleek and Puri (1985), and Hallin and Puri (1994). The results developed there (which include asymptotic linearity property) hold for unbounded square-integrable scores from which one can extract central sequence (analogous to $\boldsymbol{S}_{\phi}(\boldsymbol{\beta})$) as the leading term in the asymptotic distribution of the asymptotically optimal R-estimator. Alternatively, through the ranked-residual method of Kreiss (1990), the central sequences based on ranked-residuals can be inserted into the traditional one-step estimators discussed in Kreiss (1987) for the construction of asymptotically optimal R-estimator.

3. SOME TECHNICAL RESULTS

This section contains the statements of some technical lemmas which are used heavily in the proofs of the AUL of linear rank statistics. In the sequel, n_0 denotes a large positive integer; M denotes a large positive number; and $\mathscr{Z}$ denotes the set of all integers. Let $\{e_i; i \in \mathscr{Z}\}$ be an i.i.d. sequence of continuous random variables with $Ee_1 = 0$, $Var(e_1) = \sigma_e^2 < \infty$ and let $R_{n,i}$ denote the rank of e_i among $\{e_j; 1 \leqslant j \leqslant n\}$. For each $n \geqslant n_0$, let $\{\eta_{n,j}; j \geqslant 0\}$ be a sequence of real numbers satisfying

(3.1)

$$\sup\left\{n^{-1}\sum_{j=0}^{\infty} j\eta_{n,j}^2; n \geqslant n_0\right\} < \infty \qquad \text{and} \qquad \sup\left\{\sum_{j=0}^{\infty} |\eta_{n,j}|; n \geqslant n_0\right\} < \infty.$$

For $i \in \mathscr{Z}$, define $V_{n,i} = \sum_{j=0}^{\infty} \eta_{n,j} e_{i-j}$. The following lemma states the precise rate of convergence (in probability to zero) of the randomly weighted (by $V_{n,i}$'s) sum of the functions of ranks.

LEMMA 3.1. *For $n \geqslant 1$, $k \geqslant 1$, let b_n^k: $\{1, 2, \ldots, n\} \to \mathbb{R}^1$ be functions with*

(3.2)

$$\limsup_{k\to\infty}\limsup_{n\to\infty} E[b_n^k(R_{n,1})]^2 = \limsup_{k\to\infty}\limsup_{n\to\infty} n^{-1}\sum_{i=1}^{n} \{b_n^k(i)\}^2 = 0.$$

Then, for each fixed $l \geqslant 1$ and $\epsilon > 0$,

$$\limsup_{k\to\infty}\limsup_{n\to\infty} P\left[\left|n^{-1/2}\sum_{i=1}^{n} V_{n,i-l} b_n^k(R_{n,i})\right| > \epsilon\right] = 0. \tag{3.3}$$

The following lemma states the conditions under which (3.1) is satisfied. Here for $r > 0$, $D(0, r) = \{x \in \mathscr{C}; |x| \leqslant r\}$. A variant of this lemma can also be found in Kreiss (1987, Lemma 6.1).

LEMMA 3.2. *Let $B(x) = 1 - b_1 x - \cdots - b_p x^p$ $(b_p \neq 0)$ be a polynomial such that $B(x) \neq 0$, $\forall x \in D(0, 1)$. Then,*

(i) *$\exists \epsilon > 0$ such that $B(x) \neq 0$, $\forall x \in D(0, 1+\epsilon)$ and*

(ii) *$\exists \{\eta_j := \eta_j(b_1, \ldots, b_p); j \geqslant 0\}$ such that*

$$1/B(x) = \sum_{j=0}^{\infty} \eta_j x^j, \qquad x \in D(0, 1+\epsilon).$$

Let $\{\mathbf{t}_n = [t_{n1}, ..., t_{np}]'\}$ *be a sequence of vectors converging to zero and define* $B_n(x) = 1 - (b_1 + t_{n1})\, x - \cdots - (b_p + t_{np})\, x^p$. *Then,*

(iii) $\exists n_0$ *such that* $\forall n \geqslant n_0$, B_n *is a polynomial of degree* p *such that* $B_n(x) \neq 0$, $\forall x \in D(0, 1+\epsilon/2)$ *and*

(iv) $\forall n \geqslant n_0$, $\exists \{\eta_{n,j} := \eta_{n,j}(b_1, ..., b_p); j \geqslant 0\}$ *such that*

$$1/B_n(x) = \sum_{j=0}^{\infty} \eta_{n,j} x^j, \qquad x \in D(0, 1+\epsilon/4),$$

where for some $(M, c) \in (0, \infty) \times (0, 1)$,

$$|\eta_{nj}| \leqslant M c^j, \qquad j \geqslant 0. \tag{3.4}$$

Consequently, (3.1) *is satisfied.*

The next lemma is used in Proposition 4.2. It gives the rate of convergence (in probability to zero) of the randomly weighted (by $V_{n,i}$'s) sum of the functions of i.i.d. random variables.

LEMMA 3.3. *For* $n \geqslant 1$, $k \geqslant 1$, *let* $c_n^k\colon \mathbb{R}^1 \to \mathbb{R}^1$ *be functions with*

$$\limsup_{k\to\infty} \limsup_{n\to\infty} Var[c_n^k(e_1)] = 0. \tag{3.5}$$

Then, for each fixed $l \geqslant 1$ *and* $\epsilon > 0$,

$$\limsup_{k\to\infty} \limsup_{n\to\infty} P\left[\left| n^{-1/2} \sum_{i=1}^{n} V_{n,i-l}\{c_n^k(e_i) - E[c_n^k(e_1)]\}\right| > \epsilon\right] = 0. \tag{3.6}$$

The concept of the contiguity of probability measures, introduced by Professor Le Cam (1960) plays a basic role in the development of asymptotic distribution theory of rank statistics. Recall from Le Cam and Yang (1990, Section 3) and Jurečková (1969, Lemma 3.5) that a sequence of probability measures $\{G_n\}$ is contiguous to another sequence of probability measures $\{F_n\}$ if for any $\epsilon > 0$, there is a $\delta = \delta(\epsilon) > 0$ and $n_0 = n_0(\epsilon)$ such that for all $n \geqslant n_0$, $F_n[A] < \delta$ implies that $G_n[A] < \epsilon$. The following lemma states that the sequences of contiguous probability measures induce similar property on double sequence of sets.

LEMMA 3.4. *Let* $\{G_n\}$ *be contiguous to* $\{F_n\}$. *Let* $\{A_n^k\}$ *is a (double) sequence of sets such that*

$$\limsup_{k\to\infty} \limsup_{n\to\infty} F_n[A_n^k] = 0. \tag{3.7}$$

Then

$$\limsup_{k\to\infty}\ \limsup_{n\to\infty}\ G_n[A_n^k]=0.$$

4. ASYMPTOTIC UNIFORM LINEARITY OF RANK STATISTIC

This section contains the main results of the paper on the asymptotics of R-estimators. Let $P_{n,\boldsymbol{\theta}}$ denote the underlying stationary probability distribution of $\{X_{1-p}, ..., X_n\}$ when the true parameter is $\boldsymbol{\theta}=(\theta_1, ..., \theta_p)'$, i.e., when $X_i=\boldsymbol{\theta}'\boldsymbol{Y}_{i-1}+\varepsilon_i$, $1\leqslant i\leqslant n$, where $\{\varepsilon_i,\ 1\leqslant i\leqslant n\}$ are i.i.d. with pdf f, mean zero and variance σ^2 and the modulus of all roots of the pth degree polynomial $1-\theta_1 x-\cdots-\theta_p x^p$ are greater than unity. Under $P_{n,\boldsymbol{\theta}}$, there exist a unique sequence of real numbers $\{\psi_j(\boldsymbol{\theta}); j\geqslant 0\}$ and random variables $\{\varepsilon_k, -\infty<k\leqslant 0\}$ such that $\{\varepsilon_i, -\infty<i\leqslant n\}$ are i.i.d. and in the L^2-approximation sense

$$X_i=\sum_{j=0}^{\infty}\psi_j(\boldsymbol{\theta})\,\varepsilon_{i-j},\qquad 1-p\leqslant i\leqslant n, \tag{4.1}$$

where

$$1/(1-\theta_1 x-\cdots-\theta_p x^p)=\sum_{j=0}^{\infty}\psi_j(\boldsymbol{\theta})\,x^j,\qquad x\in\mathscr{C},\ |x|<1.$$

See Brockwell and Davis (1996, Theorem 3.1.1) for more on it. Also, the stationarity of $\{X_i\}$'s under $P_{n,\boldsymbol{\beta}}$ implies that the sequence of random matrices $n^{-1}\sum_{i=1}^{n}(Y_{i-1}-\bar{Y})(Y_{i-1}-\bar{Y})'$ converges in probability to a positive definite matrix $\boldsymbol{\Sigma}$, whose (i,j)th entry is given by $\sigma_{i,j}=Cov(X_i, X_j)$, $1\leqslant i, j\leqslant p$.

Recall from Kreiss (1987, Corollary 3.2) that the sequence of probability measures $\{P_{n,\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t}}\}$ is contiguous to the sequence of probability measures $\{P_{n,\boldsymbol{\beta}}\}$ and vice versa for each $\boldsymbol{t}\in\mathbb{R}^p$, under the following assumptions (B.3)–(B.5) [together with (B.1), (B.2)].

(B.3) The error pdf f is nonzero and absolutely continuous with finite Fisher information for location.

(B.4) For every $0\leqslant k\leqslant n$, the joint distribution of $(X_{1-p}, ..., X_0, ..., X_k)$ possesses a nowhere-vanishing Lebesgue density $g_k(\cdot,\boldsymbol{\beta})$.

(B.5) The stochastic process $g_0(X_{1-p}, ..., X_0, \cdot)$ is continuous in probability at the true parameter $\boldsymbol{\beta}$.

Notation. Let $\{\boldsymbol{U}=\boldsymbol{U}(n,k); n\geqslant 1, k\geqslant 1\}$ be a (double) sequence of random vectors. We write $\boldsymbol{U}=o_p^k(1)$, if $\limsup_{n\to\infty} P_{n,\boldsymbol{\beta}}[\|\boldsymbol{U}(n,k)\|>\epsilon]=0$

for every $k \geqslant 1$ and $\epsilon > 0$. We write $\boldsymbol{U} = o_p(1)$, if $\limsup_{k\to\infty} \limsup_{n\to\infty} P_{n,\beta}[\|\boldsymbol{U}(n,k)\| > \epsilon] = 0$ for every $\epsilon > 0$. Note that when $\boldsymbol{U}$ is a sequence in n only, $\boldsymbol{U} = o_p(1)$ means $\limsup_{n\to\infty} P_{n,\beta}[\|\boldsymbol{U}(n)\| > \epsilon] = 0$. For the boundedness in probability, we use notations $O_p(1)$ and $O_p^k(1)$ in obvious fashion. For $b > 0$, let $\mathcal{N}_b$ denote the set $\{\boldsymbol{t} \in \mathbb{R}^p;\ \|\boldsymbol{t}\| \leqslant b\}$.

The following theorem gives the AUL of the rank statistic $\boldsymbol{S}_\varphi(\boldsymbol{t})$.

THEOREM 4.1. *Let $\varphi \in \mathcal{F}$ and assumptions* (B.1)–(B.5) *hold. Then for all $b > 0$,*

$$\sup\left\{\left\|\boldsymbol{S}_\varphi(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t})-\boldsymbol{S}_\varphi(\boldsymbol{\beta})+\int f\,d\varphi(F)\,\boldsymbol{\Sigma}\boldsymbol{t}\right\|;\ \boldsymbol{t}\in\mathcal{N}_b\right\}=o_p(1). \tag{4.2}$$

Proof. Define $\mathcal{F}_c \subset \mathcal{F}$ as follows.

$$\mathcal{F}_c = \{\varphi_c;\ \varphi_c : [0,1] \to \mathbb{R}^1 \text{ is non-decreasing, right-continuous and bounded on } [0,1]\}.$$

Assertion (4.2) was proved in Koul and Ossiander (1994, Theorem 1.2 and Remark 1.1) for functions in $\mathcal{F}_c$. To extend (4.2) to functions in $\mathcal{F}$, we first show the convergence in (4.2) for each fixed $\boldsymbol{t} = [t_1, \cdots, t_p]' \in \mathcal{N}_b$. Let $\{\varphi_c^k, k \geqslant 1\}$ be a sequence of functions in $\mathcal{F}_c$ defined by

$$\varphi_c^k(u)=\sum_{i=1}^{k-1}\varphi\left(\frac{i}{k+1}\right)I\left(\frac{i-1}{k}\leqslant u<\frac{i}{k}\right)+\varphi\left(\frac{k}{k+1}\right)I\left(\frac{k-1}{k}\leqslant u\leqslant 1\right). \tag{4.3}$$

By Hájek, Šidák and Sen (1999) (Lemma 1, Section 6.1.6),

$$\lim_{k\to\infty}\int_0^1\{\varphi_c^k(u)-\varphi(u)\}^2\,du=0. \tag{4.4}$$

Writing $\boldsymbol{S}_c^k$ for $\boldsymbol{S}_{\varphi_c^k}$, we get

$$\begin{aligned}
&\left\|\boldsymbol{S}_\varphi(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t})-\boldsymbol{S}_\varphi(\boldsymbol{\beta})+\int f\,d\varphi(F)\,\boldsymbol{\Sigma}\boldsymbol{t}\right\|\\
&\quad\leqslant\|\boldsymbol{S}_\varphi(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t})-\boldsymbol{S}_c^k(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t})\|+\|\boldsymbol{S}_\varphi(\boldsymbol{\beta})-\boldsymbol{S}_c^k(\boldsymbol{\beta})\|\\
&\qquad+\left\|\int f\,d[\varphi(F)-\varphi_c^k(F)]\,\boldsymbol{\Sigma}\boldsymbol{t}\right\|\\
&\qquad+\left\|\boldsymbol{S}_c^k(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t})-\boldsymbol{S}_c^k(\boldsymbol{\beta})+\int f\,d\varphi_c^k(F)\,\boldsymbol{\Sigma}\boldsymbol{t}\right\|\\
&\quad=\|\boldsymbol{T}_1(n,k)\|+\|\boldsymbol{T}_2(n,k)\|+\|\boldsymbol{T}_3(k)\|+\|\boldsymbol{T}_4(n,k)\|,\ \text{say.}
\end{aligned}$$

By Koul and Ossiander (1994, Theorem 1.2), $\boldsymbol{T}_4(n, k) = o_p^k(1)$ for every $k \geqslant 1$. Also, integration by parts, the finiteness of the Fisher information in assumption (B.3), the choice of φ_c^k in (4.3) and (4.4) imply that $\lim_{k\to\infty} \|\boldsymbol{T}_3(k)\| = 0$. Using Lemmas 3.1, 3.2 and 3.4, we now show that

$$\boldsymbol{T}_1(n, k) = o_p(1). \tag{4.5}$$

Let $\boldsymbol{\beta}_n = \boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}$. Note that

$$\begin{aligned}\boldsymbol{T}_1(n, k) &= n^{-1/2} \sum_{i=1}^{n} \boldsymbol{Y}_{i-1} \left\{\varphi\left(\frac{R_{i\beta_n}}{n+1}\right) - \varphi_c^k\left(\frac{R_{i\beta_n}}{n+1}\right)\right\} \\ &\quad - n^{\frac{1}{2}}\, \bar{\boldsymbol{Y}} n^{-1} \sum_{i=1}^{n} \left\{\varphi\left(\frac{i}{n+1}\right) - \varphi_c^k = \left(\frac{i}{n+1}\right)\right\} \\ &= \boldsymbol{T}_{11}(n, k) - \boldsymbol{T}_{12}(n, k), \text{ say.}\end{aligned}$$

By Brockwell and Davis (1996, Theorem 7.1.2), $\|n^{\frac{1}{2}}\, \bar{\boldsymbol{Y}}\| = O_p(1)$. Also,

$$\begin{aligned}&\limsup_{k\to\infty}\, \limsup_{n\to\infty}\, n^{-1} \sum_{i=1}^{n} \left\{\varphi\left(\frac{i}{n+1}\right) - \varphi_c^k\left(\frac{i}{n+1}\right)\right\}^2 \\ &\quad = \limsup_{k\to\infty} \int_0^1 \{\varphi_c^k(u) - \varphi(u)\}^2 du = 0.\end{aligned} \tag{4.6}$$

Therefore, $\boldsymbol{T}_{12}(n, k) = o_p(1)$.

To prove $\boldsymbol{T}_{11}(n, k) = o_p(1)$, note that by the contiguity of $\{P_{n,\beta}\}$ to $\{P_{n,\beta_n}\}$ and Lemma 3.4, it suffices to show that for every $\epsilon > 0$

$$\limsup_{k\to\infty}\, \limsup_{n\to\infty}\, P_{n,\beta_n}\left[\left\| n^{-1/2} \sum_{i=1}^{n} \boldsymbol{Y}_{i-1} h_n^k\{\text{Rank}(X_i - \boldsymbol{\beta}_n' \boldsymbol{Y}_{i-1})\}\right\| > \epsilon\right] = 0,$$

where $h_n^k(i) = \varphi(\frac{i}{n+1}) - \varphi_c^k(\frac{i}{n+1})$. Since, under P_{n,β_n}, $X_i - \boldsymbol{\beta}_n' \boldsymbol{Y}_{i-1} \equiv \varepsilon_i$, it remains to show that

$$\limsup_{k\to\infty}\, \limsup_{n\to\infty}\, P_{n,\beta_n}\left[\left\| n^{-1/2} \sum_{i=1}^{n} \boldsymbol{Y}_{i-1} h_n^k\{\text{Rank}(\varepsilon_i)\}\right\| > \epsilon\right] = 0.$$

Fix an l such that $1 \leqslant l \leqslant p$ and we show that

$$\limsup_{k\to\infty}\, \limsup_{n\to\infty}\, P_{n,\beta_n}\left[\left| n^{-1/2} \sum_{i=1}^{n} X_{i-l} h_n^k\{\text{Rank}(\varepsilon_i)\}\right| > \epsilon\right] = 0. \tag{4.7}$$

Note that similar to (4.1), under P_{n,β_n},

$$X_i = \sum_{j=0}^{\infty} \psi_{n,j}\varepsilon_{i-j}, \tag{4.8}$$

where by Lemma 3.2(iv), $\{\psi_{n,j};\, j \geqslant 0\}$ satisfy

$$1/\{1-(\beta_1+n^{-1/2}t_1)\,x-\cdots-(\beta_p+n^{-1/2}t_p)\,x^p\} = \sum_{j=0}^{\infty} \psi_{n,j}x^j.$$

Next we apply Lemma 3.1 with $e_i = \varepsilon_i$, $\eta_{n,j} = \psi_{n,j}$ [since $|\psi_{n,j}| \leqslant Mc^j$ implies (3.1)], $V_{n,i} = X_i$ (of (4.8)) and $b_n^k(i) = h_n^k(i)$ [since (4.6) implies (3.2)]. From the conclusion (3.3), we obtain (4.7) and this completes the proof of (4.5). The proof of $T_2(n,k) = o_p(1)$ follows from (4.5) by specializing to $\boldsymbol{t} = 0$.

Consequently, the pointwise convergence in (4.2) for functions in $\mathscr{F}$ follows. The uniform convergence in (4.2) over $\mathcal{N}_b$ follows from the pointwise convergence in $\boldsymbol{t}$ and the convexity of the function $\boldsymbol{t} \to \sum_{i=1}^{n} (X_i - \boldsymbol{t}'\boldsymbol{Y}_{i-1})\,\varphi(R_{it}/(n+1))$ (proved in Lemmas 7.3b.1 and 7.3b.2 of Koul (1992)), along the lines of Jurečková and Sen (1996, Lemma 6.6.1), which is adapted from Heiler and Willers (1988).

The following proposition gives the asymptotic representations of R-estimators.

PROPOSITION 4.1. *Under the assumptions of Theorem* 4.1, *there exists a sequence of* $\{\hat{\boldsymbol{\beta}}_\varphi\}$ *such that*

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_\varphi - \boldsymbol{\beta}) = O_p(1). \tag{4.9}$$

Moreover,

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_\varphi - \boldsymbol{\beta}) = \left(\int f\,d\varphi(F)\right)^{-1} \Sigma^{-1}\boldsymbol{S}_\varphi(\boldsymbol{\beta}) + o_p(1). \tag{4.10}$$

Proof. The proof of (4.9) follows using an argument similar to Koul (1992, Lemma 4.4.2) and Jurečková and Sen (1996, Displays 6.6.33–6.6.35). The proof of (4.10) follows from Theorem 4.1 by substituting $\boldsymbol{t} = n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_\varphi - \boldsymbol{\beta})$ (which by (4.9) is bounded in probability) in (4.2). The details are omitted for brevity.

In order to get the asymptotic normality of the R-estimator $\hat{\boldsymbol{\beta}}_\varphi$, we need to establish the same for

$$\boldsymbol{S}_\varphi(\boldsymbol{\beta}) = n^{-1/2} \sum_{i=1}^{n} (\boldsymbol{Y}_{i-1} - \bar{\boldsymbol{Y}})\,\varphi\left(\frac{R_i}{n+1}\right).$$

But, this is a randomly weighted sum of rank scores. Moreover, the random weights $\{Y_{i-1}-\bar{Y};\, 1 \leqslant i \leqslant n\}$ as well as $\{R_1, \ldots, R_n\}$ are dependent. In (4.11) of Proposition 4.2 below, we first reduce $\boldsymbol{S}_\varphi(\boldsymbol{\beta})$ to a randomly weighted sum of independent random variables defined by

$$\begin{aligned}\hat{\boldsymbol{S}}_\varphi &= n^{-1/2}\sum_{i=1}^{n}(\boldsymbol{Y}_{i-1}-\bar{\boldsymbol{Y}})\,\varphi\,(F(\varepsilon_i))\\ &= n^{-1/2}\sum_{i=1}^{n}(\boldsymbol{Y}_{i-1}-\bar{\boldsymbol{Y}})\{\varphi(F(\varepsilon_i))-E[\varphi(F(\varepsilon_1))]\}.\end{aligned}$$

Then we establish the asymptotic normality of $\hat{\boldsymbol{S}}_\varphi$ by using multivariate martingale central limit theorem on the vector of martingale diffrences

$$\hat{\boldsymbol{S}}^*_\varphi = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Y}_{i-1}\{\varphi(F(\varepsilon_i))-E[\varphi(F(\varepsilon_1))]\}.$$

Proposition 4.2 below was proved in Koul and Ossiander (1994, Theorem 1.2, Remark 1.1 and Lemma 1.2) when the function φ is in $\mathscr{F}_c$. Here '$\Rightarrow$' denotes the convergence in distribution.

PROPOSITION 4.2. *Suppose that* (B.1) *and* (B.2) *hold. Then*

$$\boldsymbol{S}_\varphi(\boldsymbol{\beta})-\hat{\boldsymbol{S}}_\varphi = o_p(1). \tag{4.11}$$

Moreover

$$\hat{\boldsymbol{S}}_\varphi \Rightarrow N_p[0, \sigma^2_\varphi\,\Sigma], \tag{4.12}$$

where $\sigma^2_\varphi = Var[\varphi(F(\varepsilon_1))]$. *Hence under the assumptions of Theorem* 4.1

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_\varphi-\boldsymbol{\beta}) \Rightarrow N_p\left[0, \left(\int f\,d\varphi(F)\right)^{-2}\sigma^2_\varphi\,\Sigma^{-1}\right]. \tag{4.13}$$

Proof. In the following, we continue to use the notations of Theorem 4.1. Writing $\hat{\boldsymbol{S}}^k_c$ for $\hat{\boldsymbol{S}}_{\varphi^k_c}$, we get

$$\begin{aligned}&\|\boldsymbol{S}_\varphi(\boldsymbol{\beta})-\hat{\boldsymbol{S}}_\varphi\|\\ &\quad\leqslant \|\boldsymbol{S}_\varphi(\boldsymbol{\beta})-\boldsymbol{S}^k_c(\boldsymbol{\beta})\|+\|\boldsymbol{S}^k_c(\boldsymbol{\beta})-\hat{\boldsymbol{S}}^k_c\|+\|\hat{\boldsymbol{S}}^k_c-\hat{\boldsymbol{S}}_\varphi\|\\ &\quad= \|\boldsymbol{T}_5(n,k)\|+\|\boldsymbol{T}_6(n,k)\|+\|\boldsymbol{T}_7(n,k)\|, \qquad \text{say.}\end{aligned}$$

Since $T_5(n,k)=T_2(n,k)$, we get $T_5(n,k)=o_p(1)$. By Koul and Ossiander (1994, Theorem 1.2), for every $k\geqslant 1$, $T_6(n,k)=o_p^k(1)$. We next show that $T_7(n,k)=o_p(1)$. Write

$$\begin{aligned}
T_7(n,k) &\\
&= n^{-1/2}\sum_{i=1}^{n} Y_{i-1}\{\varphi_c^k(F(\varepsilon_i))-E[\varphi_c^k(F(\varepsilon_1))]-\varphi(F(\varepsilon_i))+E[\varphi(F(\varepsilon_1))]\}\\
&\quad -\bar{Y}n^{-1/2}\sum_{i=1}^{n}\{\varphi_c^k(F(\varepsilon_i))-E[\varphi_c^k(F(\varepsilon_1))]-\varphi(F(\varepsilon_i))+E[\varphi(F(\varepsilon_1))]\}\\
&= T_{71}(n,k)-T_{72}(n,k), \text{ say.}
\end{aligned}$$

Using Lemma 3.3 with $c_n^k(x)=\varphi_c^k(F(x))-\varphi(F(x))$, $T_{71}(n,k)=o_p(1)$. Since

$$n^{-1/2}\sum_{i=1}^{n}\{\varphi_c^k(F(\varepsilon_i))-E[\varphi_c^k(F(\varepsilon_1))]-\varphi(F(\varepsilon_i))+E[\varphi(F(\varepsilon_1))]\}=O_p^k(1)$$

and $\bar{Y}=o_p(1)$, we get $T_{72}(n,k)=o_p(1)$. Thus (4.11) is proved. To prove (4.12), note that

$$\hat{S}_\varphi-\hat{S}_\varphi^*=-\bar{Y}n^{-1/2}\sum_{i=1}^{n}\{\varphi(F(\varepsilon_i))-E[\varphi(F(\varepsilon_1))]\}=o_p(1)\times O_p(1)=o_p(1),$$

and $\hat{S}_\varphi^*$, being a sum of martingale difference arrays, is asymptotically normal by Corollary 3.1 of Hall and Heyde (1980). The limiting dispersion matrix is obtained by noting that the sum of the conditional dispersions $n^{-1}\sum_{i=1}^{n} Y_{i-1}Y_{i-1}'E\{\varphi(F(\varepsilon_i))-E[\varphi(F(\varepsilon_1))]\}^2$ converges in probability to $\sigma_\varphi^2\Sigma$.

Remark 4.1. From the expression of σ_ϕ^2 in (4.12) and (4.13), it follows readily that under the innovation density f, the R-estimator corresponding to the score function $\phi(u)=\gamma\{F^{-1}(u)\}$, where $\gamma(x)=-f'(x)/f(x)$ and $F^{-1}(u)$ is the u-th error-quantile, is asymptotically optimal in estimating β. Moreover, comparing $(\int f\,d\varphi(F))^{-2}\sigma_\phi^2$ and σ^2, it follows that the classical Chernoff-Savage type result holds even in the autoregressive models. In other words, the R-estimator of the autoregressive parameters based on the normal score function is not only asymptotically efficient at the Gaussian white noise, but also is superior to the standard correlogram-based estimators at other error densities.

APPENDIX

Proof of Lemma 3.1. We will show (3.3) only for the case $l = 1$, without loss of generality. For technical convenience, we only show that

$$n^{-1/2} \sum_{i=2}^{n} V_{n,i-1} b_n^k(R_{n,i}) = o_p(1). \tag{A1}$$

By definition, we have for $i \geqslant 2$,

$$V_{n,i-1} = \sum_{j=i-1}^{\infty} \eta_{n,j} e_{i-1-j} + \sum_{j=0}^{i-2} \eta_{n,j} e_{i-1-j}.$$

Thus, (A.1) follows from

$$n^{-\frac{1}{2}} \sum_{i=2}^{n} \sum_{j=i-1}^{\infty} \eta_{n,j} e_{i-1-j} b_n^k(R_{n,i}) = o_p(1),$$

and

$$n^{-\frac{1}{2}} \sum_{i=2}^{n} \sum_{j=0}^{i-2} \eta_{n,j} e_{i-1-j} b_n^k(R_{n,i}) = o_p(1).$$

But, by the Chebychev inequality, these are implied by

$$\limsup_{k \to \infty} \limsup_{n \to \infty} n^{-1} E \left[\sum_{i=2}^{n} \sum_{j=i-1}^{\infty} \eta_{n,j} e_{i-1-j} b_n^k(R_{n,i}) \right]^2 = 0 \tag{A2}$$

and

$$\limsup_{k \to \infty} \limsup_{n \to \infty} n^{-1} E \left[\sum_{i=2}^{n} \sum_{j=0}^{i-2} \eta_{n,j} e_{i-1-j} b_n^k(R_{n,i}) \right]^2 = 0. \tag{A3}$$

To prove (A.2), note that

$$\begin{aligned} &\sum_{i=2}^{n} \sum_{j=i-1}^{\infty} \eta_{n,j} e_{i-1-j} b_n^k(R_{n,i}) \\ &\quad = \sum_{i=2}^{n} \sum_{l=-\infty}^{0} \eta_{n,i-1-l} e_l b_n^k(R_{n,i}) = \sum_{l=-\infty}^{0} e_l \sum_{i=2}^{n} \eta_{n,i-1-l} b_n^k(R_{n,i}), \end{aligned}$$

and $\{e_l, l \leqslant 0\}$ and $\{R_{n1}, ..., R_{nn}\}$ are independent. Consequently,

$$E\left[\sum_{i=2}^{n} \sum_{j=i-1}^{\infty} \eta_{n,j} e_{i-1-j} b_n^k(R_{n,i})\right]^2 = Var\left[\sum_{l=-\infty}^{0} e_l \sum_{i=2}^{n} \eta_{n,i-1-l} b_n^k(R_{n,i})\right]$$

$$= \sigma^2 \sum_{l=-\infty}^{0} Var\left[\sum_{i=2}^{n} \eta_{n,i-1-l} b_n^k(R_{n,i})\right]$$

$$\leqslant \sigma^2 \sum_{l=-\infty}^{0} (n-1)^{-1} \sum_{i=2}^{n} \eta_{n,i-1-l}^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2$$

$$\leqslant \sigma^2 (n-1)^{-1} \sum_{l=-\infty}^{0} \sum_{i=2}^{n} \eta_{n,i-1-l}^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2$$

$$\leqslant \sigma^2 (n-1)^{-1} \left(\sum_{j=0}^{\infty} (j+1)\, \eta_{n,j}^2\right) \sum_{i=1}^{n} \{b_n^k(i)\}^2,$$

where the first inequality follows from the variance formula for linear rank statistic (c.f. Hájek, Šidák and Sen (1999, Theorem 3, Section 3.3)) and the last inequality follows from $\sum_{l=-\infty}^{0} \sum_{i=2}^{n} \eta_{n,i-1-l}^2 \leqslant \sum_{j=0}^{\infty} (j+1)\, \eta_{n,j}^2$, obtained by interchanging the two sums. Hence (A.2) follows from assumptions (3.1) and (3.2).

To prove (A.3), we have

$$E\left[\sum_{i=2}^{n} \sum_{j=0}^{i-2} \eta_{n,j} e_{i-1-j} b_n^k(R_{n,i})\right]^2$$

$$= E\left[\sum_{j=0}^{n-2} \eta_{n,j} \sum_{i=j+2}^{n} e_{i-1-j} b_n^k(R_{n,i})\right]^2$$

$$\leqslant \sum_{j=0}^{n-2} |\eta_{n,j}| \sum_{j=0}^{n-2} |\eta_{n,j}|\, E\left[\sum_{i=j+2}^{n} e_{i-1-j} b_n^k(R_{n,i})\right]^2$$

$$\leqslant \left(\sum_{j=0}^{\infty} |\eta_{n,j}|\right)^2 \max_{0 \leqslant j \leqslant n-2} E\left[\sum_{i=j+2}^{n} e_{i-1-j} b_n^k(R_{n,i})\right]^2.$$

Since $\sup\{\sum_{j=0}^{\infty} |\eta_{n,j}|; n \geqslant n_0\} < \infty$, assertion (A.3) will follow from

$$\text{(A4)} \qquad \limsup_{k\to\infty} \limsup_{n\to\infty} n^{-1} \max_{0 \leqslant j \leqslant n-2} E\left[\sum_{i=j+2}^{n} e_{i-1-j} b_n^k(R_{n,i})\right]^2 = 0.$$

To prove (A.4), note that for each fixed j with $0 \leqslant j \leqslant n-2$, we have

(A5) $$E\left[\sum_{i=j+2}^{n} e_{i-1-j} b_n^k(R_{n,i})\right]^2$$

$$= \sum_{i=j+2}^{n} E[e_{i-1-j}^2\{b_n^k(R_{n,i})\}^2]$$

$$+2 \sum_{j+2 \leqslant i_1 < i_2(=i_1+1+j) \leqslant n} E[e_{i_1-1-j}e_{i_2-1-j}b_n^k(R_{n,i_1})\, b_n^k(R_{n,i_2})]$$

$$+2 \sum_{j+2 \leqslant i_1 < i_2(\neq i_1+1+j) \leqslant n} E[e_{i_1-1-j}e_{i_2-1-j}b_n^k(R_{n,i_1})\, b_n^k(R_{n,i_2})]$$

$$= I_1(j) + 2I_2(j) + 2I_3(j), \text{ say.}$$

First we bound $I_1(j)$ and $I_2(j)$. Note that for $1 \leqslant k < i$, $(e_k, R_{n,i})$ and $(e_1, R_{n,i})$ are identically distributed. Therefore, for $i \geqslant 2$ and $0 \leqslant j \leqslant n-2$, $(e_{i-1-j}, R_{n,i})$ and $(e_1, R_{n,i})$ are identically distributed. Hence,

$$I_1(j) = \sum_{i=j+2}^{n} E[e_1^2\{b_n^k(R_{n,i})\}^2] \leqslant Ee_1^2 \sum_{i=1}^{n} \{b_n^k(R_{n,i})\}^2 = \sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2.$$

Also, $|I_2(j)| = |\sum_{i=j+2}^{n-1-j} E[e_{i-1-j}e_i\, b_n^k(R_{n,i})\, b_n^k(R_{n,i+1+j})]|$, which by the Cauchy–Schwartz inequality is bounded by

$$\left(\sum_{i=j+2}^{n-1-j} E[e_{i-1-j}^2\{b_n^k(R_{n,i})\}^2] \sum_{i=j+2}^{n-1-j} E[e_i^2\{b_n^k(R_{n,i+1+j})\}^2]\right)^{1/2}$$

$$\leqslant \sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2.$$

Finally, more delicate argument is needed to bound $I_3(j)$. Note that

$$I_3(j) = \sum_{j+2 \leqslant i_1 < i_2(\neq i_1+1+j) \leqslant n} E[e_{i_1-1-j}e_{i_2-1-j}b_n^k(R_{n,i_1})\, b_n^k(R_{n,i_2})]$$

$$= N(j)\, E[e_1 e_2 b_n^k(R_{n,3})\, b_n^k(R_{n,4})],$$

where $N(j)$ is the cardinality of the set

$$\{(i_1, i_2);\, j+2 \leqslant i_1 < i_2(\neq i_1+1+j) \leqslant n\}.$$

When $n \geqslant 2j+2$, $N(j) = (n-j-3)(n-j-2)/2 + j$ and when $n < 2j+2$, $N(j) = (n-j-2)(n-j-1)/2$. In both cases, $N(j) \leqslant n^2/2$.

Next, we bound $E[e_1 e_2 b_n^k(R_{n,3})\, b_n^k(R_{n,4})]$. Denote $n(n-1)(n-2)(n-3)$ by $(n)_4$ and denote the i-th order statistic of $\{e_1, \ldots, e_n\}$ by $e_{(i)}$, $1 \leqslant i \leqslant n$. Then, using $e_i = e_{(R_{n,i})}$ (for every $1 \leqslant i \leqslant n$) and the independence among the rank and order statistics

$$
\begin{aligned}
&(n)_4\ E[e_1 e_2 b_n^k(R_{n,3})\, b_n^k(R_{n,4})] \\
&\quad = (n)_4\ E[e_{(R_{n,1})} e_{(R_{n,2})} b_n^k(R_{n,3})\, b_n^k(R_{n,4})] \\
&\quad = \sum_{i=1}^{n} \sum_{j=1(\neq i)}^{n} b_n^k(i)\, b_n^k(j) \sum_{k=1(\neq i,j)}^{n} \sum_{l=1(\neq i,j,k)}^{n} E[e_{(k)} e_{(l)}].
\end{aligned}
$$

First we evaluate the two innermost sums. Let $S_n = \sum_{i=1}^{n} e_i$. Then

$$
\begin{aligned}
&\sum_{k=1(\neq i,j)}^{n} \sum_{l=1(\neq i,j,k)}^{n} E[e_{(k)} e_{(l)}] \\
&\quad = E\left[\sum_{k=1(\neq i,j)}^{n} e_{(k)}\right]^2 - \sum_{k=1(\neq i,j)}^{n} Ee_{(k)}^2 \\
&\quad = E(S_n - e_{(i)} - e_{(j)})^2 - ES_n^2 + Ee_{(i)}^2 + Ee_{(j)}^2 \\
&\quad = 2\{Ee_{(i)}^2 + Ee_{(j)}^2 + E[e_{(i)} e_{(j)}] - E[S_n e_{(i)}] - E[S_n e_{(j)}]\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&(n)_4\ E[e_1 e_2 b_n^k(R_{n,3})\, b_n^k(R_{n,4})]/2. \\
&\quad = \sum_{i=1}^{n} \sum_{j=1(\neq i)}^{n} b_n^k(i)\, b_n^k(j)\, Ee_{(i)}^2 + \sum_{i=1}^{n} \sum_{j=1(\neq i)}^{n} b_n^k(i)\, b_n^k(j)\, Ee_{(j)}^2 \\
&\qquad + \sum_{i=1}^{n} \sum_{j=1(\neq i)}^{n} b_n^k(i)\, b_n^k(j)\, E[e_{(i)} e_{(j)}] \\
&\qquad - \sum_{i=1}^{n} \sum_{j=1(\neq i)}^{n} b_n^k(i)\, b_n^k(j)\, E[S_n e_{(i)}] - \sum_{i=1}^{n} \sum_{j=1(\neq i)}^{n} b_n^k(i)\, b_n^k(j)\, E[S_n e_{(j)}]\} \\
&\quad = J_1 + J_2 + J_3 - J_4 - J_5, \qquad \text{say.}
\end{aligned}
$$

Next we bound J_i's by using

$$
\max\{E[e_{(i)}^2];\ 1 \leqslant i \leqslant n\} \leqslant \sum_{i=1}^{n} Ee_{(i)}^2 = ES_n^2 = n\sigma^2.
$$

Towards that, by the Cauchy–Schwarz inequality,

$$
\begin{aligned}
|J_1| &\leqslant \sum_{i=1}^{n} |b_n^k(i)|\, |E[e_{(i)}^2]| \sum_{j=1}^{n} |b_n^k(j)| \\
&\leqslant \left(\sum_{i=1}^{n} \{b_n^k(i)\}^2 \sum_{i=1}^{n} E^2 e_{(i)}^2\right)^{\frac{1}{2}} \left(n \sum_{j=1}^{n} \{b_n^k(j)\}^2\right)^{\frac{1}{2}}
\end{aligned}
$$

$$\leqslant \left(\sum_{i=1}^{n} \{b_n^k(i)\}^2 n\sigma^2 \sum_{i=1}^{n} Ee_{(i)}^2 \right)^{\frac{1}{2}} \left(n \sum_{j=1}^{n} \{b_n^k(j)\}^2 \right)^{\frac{1}{2}}$$

$$= n^{\frac{3}{2}} \sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2.$$

Using similar tricks, $|J_2| \leqslant n^{\frac{3}{2}}\sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2$ and $|J_3| \leqslant n\sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2$. Next, we bound J_4 as follows.

$$|J_4| \leqslant \sum_{i=1}^{n} |b_n^k(i)| \, |E[S_n e_{(i)}]| \sum_{j=1}^{n} |b_n^k(j)|$$

$$\leqslant \left(\sum_{i=1}^{n} \{b_n^k(i)\}^2 \sum_{i=1}^{n} E^2 |S_n e_{(i)}| \right)^{\frac{1}{2}} \left(n \sum_{j=1}^{n} \{b_n^k(j)\}^2 \right)^{\frac{1}{2}}$$

$$= n^{\frac{1}{2}} \left[\sum_{i=1}^{n} \{b_n^k(i)\}^2 \right] ES_n^2$$

$$= n^{\frac{3}{2}} \sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2.$$

Similarly, $|J_5| \leqslant n^{\frac{3}{2}}\sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2$. Hence from (A.5),

$$n^{-1} \max_{0 \leqslant j \leqslant n-2} E\left[\sum_{i=j+2}^{n} e_{i-1-j} b_n^k(R_{n,i}) \right]^2$$

$$\leqslant \sigma^2 n^{-1} \sum_{i=1}^{n} \{b_n^k(i)\}^2 \left[3 + 2\frac{n_2}{(n)_4} \{n + 4n^{\frac{3}{2}}\} \right].$$

Therefore, (A.4) follows from (3.2). This also completes the proof of (A.3).

Remark A.1. An easier way to bound J_i's (good enough for proving (A.4)) is as follows. For illustration, we bound J_1 only.

$$|J_1| \leqslant \max\{E[e_{(i)}^2]; 1 \leqslant i \leqslant n\} \left(\sum_{i=1}^{n} |b_n^k(i)| \right)^2 \leqslant n^2\sigma^2 \sum_{i=1}^{n} \{b_n^k(i)\}^2.$$

Proof of Lemma 3.2. Assertions (i) and (ii) are well-known; see Brockwell and Davis (1996, Theorem 3.1.1). Assertion (iii) follows by noting that

$$\lim_{n\to\infty} \sup \{|B_n(x) - B(x)|; x \in D(0, 1+\epsilon)\} = 0.$$

To prove (iv), write $B(x)=\prod_{k=1}^{p}(1-x/\lambda_k)$, where $\{\lambda_1, ..., \lambda_p\}$ are the roots of B with $\min\{|\lambda_1|, ..., |\lambda_p|\}>1+\epsilon$. Let $\{\lambda_{n1}, ..., \lambda_{np}\}$ denote the roots of B_n. Since t_n converges to zero, by Bai (1985),

$$\lim_{n\to\infty} \min\{|\lambda_{n1}|, ..., |\lambda_{np}|\}=\min\{|\lambda_1|, ..., |\lambda_p|\}.$$

Hence, there is an n_0 such that $\forall n \geqslant n_0$,

$$\text{(A6)} \qquad \min\{|\lambda_{n1}|, ..., |\lambda_{np}|\} \geqslant 1+\epsilon/2.$$

For $x \in D(0, 1+\epsilon/4)$,

$$B_n^{-1}(x)=\prod_{k=1}^{p}(1-x/\lambda_{nk})^{-1}=\prod_{k=1}^{p}\left(\sum_{l=0}^{\infty} x^l/\lambda_{nk}^l\right)=\sum_{j=0}^{\infty}\eta_{n,j}x^j,$$

where $\eta_{n,j}=\sum_{l_i\geqslant 0;\, l_1+\cdots l_p=j}\prod_{k=1}^{p}\lambda_{nk}^{-l_k}$. By (A.6), $|\prod_{k=1}^{p}\lambda_{nk}^{-l_k}|\leqslant(1+\epsilon/2)^{-j}$. Also, the cardinality of $\{(l_1, ..., l_p); l_i\geqslant 0, l_1+\cdots l_p=j\}$ is $\binom{j-1}{p-1}\leqslant j^p$. Therefore,

$$|\eta_{n,j}|\leqslant j^p(1+\epsilon/2)^{-j}\leqslant \sup\{j^p(1+\epsilon/4)^j(1+\epsilon/2)^{-j}; j\geqslant 0\}(1+\epsilon/4)^{-j}.$$

Hence the proof of (3.4) is complete.

Proof of Lemma 3.3. Note that for any $l\geqslant 1$, $\sum_{i=1}^{n} n^{-1/2}V_{n,i-l}\{c_n^k(e_i)-E[c_n^k(e_1)]\}$ is a sum of martingale differences. Therefore,

$$E\left[n^{-1/2}\sum_{i=1}^{n}V_{i-l}\{c_n^k(e_i)-E[c_n^k(e_1)]\}\right]^2$$
$$=n^{-1}\sum_{i=1}^{n}E[V_{i-l}]^2\,E\{c_n^k(e_i)-E[c_n^k(e_1)]\}^2$$
$$=E[V_0]^2\operatorname{Var}\{c_n^k(e_1)\},$$

which converges to zero by (3.5). Hence, the proof of (3.6) is complete.

Proof of Lemma 3.4. Fix $\epsilon>0$ and let δ and $n_0(\epsilon)$ be as in the definition of contiguity given above Lemma 3.4. By the definition of limsup with respect to n, for any $k\geqslant 1$, there exists an $n(k)\geqslant n_0(\epsilon)$ such that $\forall n\geqslant n(k)$, $F_n[A_n^k]<\limsup_{n\to\infty}F_n[A_n^k]+\frac{1}{k}$. But, by (3.7), there is a $k_0(\epsilon)$ such that $\forall k\geqslant k_0(\epsilon)$, $\limsup_{k\to\infty}\limsup_{n\to\infty}F_n[A_n^k]+\frac{1}{k}<\delta$. Hence, $\forall k\geqslant k_0(\epsilon)$, there is $n_0(k)\geqslant n_0(\epsilon)$ such that $\forall n\geqslant n_0(k)\geqslant n_0(\varepsilon)$, $F_n[A_n^k]<\delta$ and consequently, $G_n[A_n^k]<\epsilon$. Since ϵ is arbitrary, the conclusion follows by taking limsups of $G_n[A_n^k]$ with respect to n, and k successively.

ACKNOWLEDGMENTS

The first author thanks Professor Hira Koul for suggesting the problem to him and for his many useful suggestions on the content and the presentation of the paper. Both authors thank the referees for their very careful reading of the paper and for their many constructive suggestions regarding the readability and the presentation of the paper.

REFERENCES

1. J. N. Adichie, Estimates of regression parameters based on rank tests, *Ann. Math. Statist.* **38** (1967), 894–904.
2. Z. D. Bai, A note on the limiting distribution of the eigenvalues of a class of random matrices, *J. of Mathematical Research and Exposition* **5** (1985), 113–118.
3. P. J. Brockwell and R. A. Davis, "Time Series: Theory and Methods," Springer-Verlag, New York, (1996).
4. H. Chernoff and I. R. Savage, Asymptotic normality and efficiency of certain nonparametric test statistics, *Ann. Math. Statist.* **29** (1958), 972–994.
5. L. Denby and D. Martin, Robust estimation of the first order autoregressive parameter, *J. Amer. Statist. Assoc.* **74** (1979), 140–146.
6. J. Hájek, Z. Šidák, and P. K. Sen, "Theory of Rank Tests," Academic Press, San Diego, 1999.
7. P. Hall and C. C. Heyde, "Martingale Limit Theory and Its Applications," Academic Press, New York, 1980.
8. M. Hallin, J. Ingenbleek, and M. Puri, Linear serial rank tests for randomness against ARMA alternatives, *Ann. Statist.* **13** (1985), 1156–1181.
9. M. Hallin and M. Puri, Aligned rank tests for linear models with autocorrelated error terms, *J. Multivariate Anal.* **50** (1994), 175–237.
10. S. Heiler and R. Willers, Asymptotic normality of R-Estimates in the linear model, *Statistics* **19** (1988), 173–184.
11. P. J. Huber, "Robust Statistics," Wiley, New York, 1981.
12. L. A. Jaeckel, Estimating regression coefficients by minimizing the dispersion of the residuals, *Ann. Math. Statist.* **43** (1972), 1449–1458.
13. J. Jurečková, Asymptotic linearity of a rank statistic in regression parameter, *Ann. Math. Statist.* **40** (1969), 1889–1900.
14. J. Jurečková, Nonparametric estimates of regression coefficients, *Ann. Math. Statist.* **42** (1971), 1328–1338.
15. J. Jurečková and P. K. Sen, "Robust Statistical Procedures: Asymptotics and Interrelations," Wiley, New York, 1996.
16. H. L. Koul, Asymptotic behavior of a class of confidence region based on rank in regression, *Ann. Math. Statist.* **42** (1971), 466–476.
17. H. L. Koul, Minimum distance estimation and goodness of fit in first-order autoregression, *Ann. Statist.* **14** (1986), 466–476.
18. H. L. Koul, A weak convergence result useful in robust autoregression, *J. Statist. Plann. Inference* **29** (1991), 291–308.
19. H. L. Koul, "Weighted Empiricals and Linear Models," Lecture Notes-Monograph Series, Vol. 21, IMS, Hayward, California, 1992.
20. H. L. Koul and A. K. Md. E. Saleh, R-estimation of the parameters of autoregressive [AR(p)] models. *Ann. Statist.* **21** (1993), 534–551.

21. H. L. Koul and M. Ossiander, Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.* **22** (1994), 540–562.
22. H. L. Koul and A. K. Md. E. Saleh, Autoregression quantiles and related rank-scores processes, *Ann. Statist.* **23** (1995), 670–689.
23. J. Kreiss, On adaptive estimation in stationary ARMA processes, *Ann. Statist.* **15** (1987), 112–133.
24. J. Kreiss, Testing linear hypotheses in autoregressions, *Ann. Statist.* **18** (1990), 1470–1482.
25. L. LeCam, Locally asymptotically normal families of distributions, *Univ. Calif. Publ. Statist.* **3** (1960), 37–98.
26. L. Le Cam and G. L. Yang, "Asymptotics in Statistics: Some Basic Concepts," Springer-Verlag, New York, 1990.
27. P. K. Sen, On a class of rank order tests for the parallelism of several regression lines, *Ann. Math. Statist.* **40** (1969), 1668–1683.